Tentamen: MVE051/MSG810 Matematisk statistik och diskret matematik, MSG820 Statistik för fysiker.
Tid och plats: Torsdag em den 15 jan, 2015, V-huset.
Examinator: Marina Axelson-Fisk. Tel: 070-2288113.
Tillåtna hjälpmedel: typgodkänd miniräknare, Beta formelsamlning.
****

Lösningar

1. $P(A) = 0.3, P(B) = 0.5$:

   (a) Disjunkta: $P(A \cap B) = 0$.

   (b) Oberoende: $P(A \cap B) = P(A)P(B) = 0.15$.

   (c) $P(A \cup B) = 0.2$:

   $$P(A \cup B) = P(A) + P(B) - P(A \cap B) \Leftrightarrow P(A \cap B) = P(A) + P(B) - P(A \cup B)$$

   så

   $$P(A \cap B) = 0.3 + 0.5 - 0.2 = 0.6$$

   Detta är inte möjligt! Snittet kan inte vara större än unionen.

   Lösning 2 (med korrekta siffror), för $P(A \cup B) = 0.6$:

   $$P(A \cap B) = 0.3 + 0.5 - 0.6 = 0.2$$

2. $X \sim \text{Bin}(n, p) \Rightarrow E[X] = np, Var(X) = np(1-p)$

   (a)

   $$\begin{cases} np & = & 5 \\ np(1-p) & = & 4 \end{cases} \Leftrightarrow \begin{cases} n & = & 25 \\ p & = & 1/5 \end{cases}$$

   (b)

   $$\begin{aligned} P(X > 3) &= 1 - P(X \leq 3) \\ &= 1 - \sum_{x=0}^{3} \binom{25}{x} 0.2^x 0.8^{25-x} \\ &= 1 - (0.8^{25} + 25 \cdot 0.2 \cdot 0.8^{24} + 300 \cdot 0.2^2 \cdot 0.8^{23} + 2300 \cdot 0.2^3 \cdot 0.8^{22}) \\ &\approx 0.77 \end{aligned}$$

3. Här använder vi Bayes formel. Vi definierar händelserna

   $$\begin{aligned} A &= \text{Zlatan gör mål} \\ B &= \text{PSG vann matchen} \end{aligned}$$ Bayes formel ger

   $$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

   Vi har från uppgiften att:

   $$\begin{aligned} P(A) &= 0.65 \\ P(A^c) &= 1 - P(A) = 0.35 \\ P(B|A) &= 0.84 \\ P(B|A^c) &= 0.52 \end{aligned}$$

Så vi får

$$P(B|A) = \frac{0.84 \cdot 0.65}{0.84 \cdot 0.65 + 0.52 \cdot 0.35} = 0.75$$

Svar: Sannolikheten är 0.75.

4. X och Y utgör alltså sidorna på en rektangel. Med andra ord har vi

$$\begin{aligned} O &= 2X + 2Y \\ A &= XY \end{aligned}$$

Vi använder att för oberoende variabler gäller $E[XY] = E[X]E[Y]$ och för lika fördelade gäller $E[X] = E[Y]$ och $\text{Var}(X) = \text{Var(Y)}$. Vi får

$$\begin{aligned} Cov(O, A) &= Cov(XY, 2X + 2Y) \\ &= E[2X^2Y + 2XY^2] - E[XY]E[2X + 2Y] \\ &= 2E[X^2]E[Y] + 2E[X]E[Y^2] - 2(E[X])^2E[Y] - 2E[X](E[Y])^2 \\ &= 4E[X](E[X^2] - (E[X])^2) \\ &= 4E[X]\text{Var}(X) \end{aligned}$$

Vi vet att X och Y är Likf$[0, 1]$, så vi vet att

$$\begin{aligned} E[X] &= E[Y] = \frac{1}{2} \\ \text{Var}(X) &= \text{Var}(Y) = \frac{1}{12} \end{aligned}$$

Sätter vi in detta i formeln får vi

$$Cov(O, A) = 4 \cdot \frac{1}{2} \cdot \frac{1}{12} = \frac{1}{6}$$

Alternativ lösning:

$$Cov(O, A) = E[OA] - E[O]E[A]$$

Obs: O och A är **inte** oberoende.

$$\begin{aligned} E[O] &= E[2X + 2Y] = 2E[X] + 2E[Y] = 2 \\ E[A] &= E[XY] = E[X]E[Y] = \frac{1}{4} \\ E[OA] &= E[(2X + 2Y)XY] \\ &= E[2X^2Y + 2XY^2] \\ &= 2E[X^2]E[Y] + 2E[X]E[Y^2] \\ &= 2(Var(X) + (E[X])^2)E[Y] + 2E[X](Var(Y) + (E[Y]^2)) \\ &= 2\left(\frac{1}{12} + \left(\frac{1}{2}\right)^2\right)\frac{1}{2} + 2\frac{1}{2}\left(\frac{1}{12} + \left(\frac{1}{2}\right)^2\right) = \frac{2}{3} \\ Cov(O, A) &= \frac{2}{3} - 2 \cdot \frac{1}{4} = \frac{1}{6} \end{aligned}$$

Svar: $Cov(O, A) = \frac{1}{6}$

5. (a) För att $f_{XY}(x, y)$ ska vara en täthetsfunktion ska den integreras till 1, dvs

$$\int\int \frac{x + 2y}{c}\, dx\, dy = 1 \Leftrightarrow \int\int (x + 2y) dx\, dy = c$$

Vi får

$$\begin{aligned}
c &= \int_0^3 \int_x^3 (x+2y)dy\,dx \\
&= \int_0^3 \left[xy+y^2\right]_x^3 \\
&= \int_0^3 (3x+9-2x^2)dx \\
&= \left[\frac{3}{2}x^2+9x-\frac{2}{3}x^3\right]_0^3 \\
&= \frac{45}{2} = 22.5
\end{aligned}$$

Svar: $c = 22.5$

Alternativ lösning: (eftersom det var lite otydligt i uppgiften om fördelning var diskret eller kontinuerlig så ger vi rätt för båda.)
För att $f_{XY}(x,y)$ ska vara en frekvensfunktion (diskret) ska den summeras till 1. Dvs

$$\sum_{x=0}^{3}\sum_{y=x}^{3}(x+2y) = c$$

Elementen i summan $(x+2y)$ för $0 \le x < y \le 3$ är

$x \backslash y$	0	1	2	3	$f_X(x)$
0	0	2	4	6	12
1	0	0	5	7	12
2	0	0	0	8	8
3	0	0	0	0	0
$f_Y(y)$	0	2	9	21	32

Svar: $c = 32$.

(b) $P(X>2) = \int_2^3 f_X(x)\,dx$ där $f_X(x)$ är X:s marginalfördelning, och

$$\begin{aligned}
f_X(x) &= \int f_{X,Y}(x,y)dy \\
&= \frac{2}{45}\int_x^3 (x+2y)dy \\
&= \frac{2}{45}\left[xy+y^2\right]_x^3 \\
&= \frac{2}{45}(3x+9-2x^2)
\end{aligned}$$

så

$$\begin{aligned}
P(X>2) &= \frac{2}{45}\int_2^3 (3x+9-2x^2)dx \\
&= \frac{2}{45}\left[\frac{3}{2}x^2+9x-\frac{2}{3}x^3\right]_2^3 \\
&= \frac{2}{45}\left(\frac{27}{2}+27-18-\left(6+18-\frac{16}{3}\right)\right) \\
&= 0.170
\end{aligned}$$

6. Vi bildar en Markovkedja med tillstånd $\{1,2,3\} = \{KTH, CTH, LTH\}$ och övergångsmatrix

$$P = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0 & 1 & 0 \\ 0.3 & 0.3 & 0.4 \end{pmatrix}$$

Vi löser ekvationssystemet

$$\left\{ \begin{array}{lcl} m_1 & = & 1 + 0.7m_1 + 0.2m_2 + 0.1m_3 \\ m_2 & = & 0 \\ m_3 & = & 1 + 0.3m_1 + 0.3m_2 + 0.4m_3 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{lcl} m_1 & = & 14/3 \\ m_2 & = & 0 \\ m_3 & = & 4 \end{array} \right.$$

Svar: 4 generationer.

7. (a) Vikten $Y \sim N(\mu, \sigma^2)$. 95% konfidensintervall för μ när σ^2 är okänd:

$$\mu = \bar{Y} \pm t_{7,0.975} \frac{s}{\sqrt{n}}$$

där

$$\begin{array}{rcl} n & = & 8 \\ t_{7,0.975} & = & 2.365 \\ s_y^2 & = & \dfrac{1}{n-1} \left(\sum y_i^2 - n\bar{y}^2 \right) = 297.43 \end{array}$$

Svar: $\mu = 31.5 \pm 2.365 \frac{17.25}{\sqrt{8}} = 31.5 \pm 14.42$.

(b) Regressionslinjen är $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ där

$$\begin{array}{rcl} \hat{\beta}_1 & = & \dfrac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = 47.21 \\ \hat{\beta}_0 & = & \bar{y} - \hat{\beta}_1 \bar{x} = -62.92 \end{array}$$

Med $x_0 = 2.3$ får vi $y_0 = -62.92 + 47.21 \cdot 2.3 = 45.66$.

(c) Vi testar hypotesen $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 > 0$ med teststatistikan

$$T = \frac{\hat{\beta}_1}{s/\sqrt{S_{xx}}} \sim t_{n-2}$$

där

$$\begin{array}{rcl} S_{xx} & = & \sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2 = 0.86 \\ s^2 & = & \dfrac{SSE}{n-2} = \dfrac{165.30}{6} = 27.55 \end{array}$$

Vi får

$$T = \frac{47.25}{\sqrt{27.55/0.86}} = 8.34$$

För ett ensidigt test slår vi upp $t_{n-2,0.99} = 3.143$, och eftersom $T > 3.143$ förkastar vi H_0. Det finns ett signifikant samband mellan vikt och längd hos alligatorer!